



Task-Modulated Cortical Representations of Natural Sound Source Categories

Hjortkjær, Jens; Kassuba, Tanja; Madsen, Kristoffer Hougaard; Skov, Martin ; Siebner, Hartwig R.

Published in:
Cerebral Cortex

Link to article, DOI:
[10.1093/cercor/bhx263](https://doi.org/10.1093/cercor/bhx263)

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Hjortkjær, J., Kassuba, T., Madsen, K. H., Skov, M., & Siebner, H. R. (2018). Task-Modulated Cortical Representations of Natural Sound Source Categories. *Cerebral Cortex*, 28(1), 295-306.
<https://doi.org/10.1093/cercor/bhx263>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

ORIGINAL ARTICLE

Task-Modulated Cortical Representations of Natural Sound Source Categories

Jens Hjortkjær^{1,2}, Tanja Kassuba³, Kristoffer H. Madsen^{1,4}, Martin Skov^{1,5} and Hartwig R. Siebner^{1,6}

¹Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre, 2650 Hvidovre, Denmark, ²Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, ³Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA, ⁴Cognitive Systems, Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, ⁵Decision Neuroscience Research Group, Copenhagen Business School, 2000 Frederiksberg, Denmark and ⁶Department of Neurology, Copenhagen University Hospital Bispebjerg, Copenhagen, 2400 København NV, Denmark

Address correspondence to Jens Hjortkjær, Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital Hvidovre, Kettegaard Allé 30, 2650 Hvidovre, Denmark. Email: jensh@drcmr.dk

Martin Skov and Hartwig R. Siebner both contributed equally to the article

Abstract

In everyday sound environments, we recognize sound sources and events by attending to relevant aspects of an acoustic input. Evidence about the cortical mechanisms involved in extracting relevant category information from natural sounds is, however, limited to speech. Here, we used functional MRI to measure cortical response patterns while human listeners categorized real-world sounds created by objects of different solid materials (glass, metal, wood) manipulated by different sound-producing actions (striking, rattling, dropping). In different sessions, subjects had to identify either material or action categories in the same sound stimuli. The sound-producing action and the material of the sound source could be decoded from multivoxel activity patterns in auditory cortex, including Heschl's gyrus and planum temporale. Importantly, decoding success depended on task relevance and category discriminability. Action categories were more accurately decoded in auditory cortex when subjects identified action information. Conversely, the material of the same sound sources was decoded with higher accuracy in the inferior frontal cortex during material identification. Representational similarity analyses indicated that both early and higher-order auditory cortex selectively enhanced spectrotemporal features relevant to the target category. Together, the results indicate a cortical selection mechanism that favors task-relevant information in the processing of nonvocal sound categories.

Key words: auditory cortex, category-based attention, fMRI, MVPA, natural sounds, object categorization, RSA

Introduction

It is commonly accepted that the primate auditory cortex is organized in a hierarchical manner, but the neural processes involved in transforming lower-level acoustic input into “auditory objects”

are still debated (Griffiths and Warren 2004; Bizley and Cohen 2013). A key question is understanding how the brain integrates acoustically different sounds into invariant categories. Functional brain mapping studies have shown that responses to

progressively more abstract “object” properties increase along an anteroventral axis in temporal cortex (Zatorre et al. 2004; Leaver and Rauschecker 2010). Category-representations in more posterior parts of nonprimary auditory cortex have been attributed to sequential information such as action categories (Giordano et al. 2012; Engel et al. 2009; Pizzamiglio et al. 2005). However, in order to function optimally in acoustically complex scenarios, the brain must not only group sounds across acoustic variation. The brain must also be able to flexibly select a category interpretation of a particular sound stimulus that is relevant in ongoing behavior. In everyday sound scenes, we rely critically on our ability to identify relevant auditory object information while ignoring irrelevant aspects. Yet, the cortical mechanisms responsible for the formation of goal-dependent representations of natural sound categories remain poorly understood.

Auditory attention is known to modulate neural responses throughout the auditory cortex. Task-related modulation of sound representations may already emerge in early auditory cortex. Animal electrophysiology studies convergently suggest that attention to a specific tone frequency sharpens receptive field selectivity of neurons in primary auditory cortex (PAC) in favor of the behaviorally relevant target (Fritz et al. 2003; Atiani et al. 2009; O’Connell et al. 2014). In human auditory cortex, enhanced responses have been demonstrated with tasks involving selective attention to a particular sound stimulus feature (Jäncke et al. 1999; Brechmann and Scheich 2004; Kauramäki et al. 2007; Paltoglou et al. 2009, 2011; Da Costa et al. 2013; Riecke et al. 2016), to a particular sound stream (Hillyard et al. 1973; Ahveninen et al. 2011), or to a spatial location (Ahveninen et al. 2006; Degerman et al. 2006). In everyday listening, however, attention usually operates at the level of object category representations. For instance, we typically attend to a “car” rather than to the pitch or loudness of the sound that the car produces (Gaver 1993; Alain and Arnott 2000). Selectively attending to specific category features of speech (e.g., specific speech content or specific talkers) is known to enhance responses in speech-sensitive regions across the superior temporal gyrus/sulcus (von Kriegstein et al. 2010; Desai et al. 2008; Andics et al. 2010; Mesgarani and Chang 2012; Bonte et al. 2014) but also in earlier regions of the auditory cortex (Kilian-Hütten et al. 2011).

At present, evidence for task-dependent processing in human auditory cortex is limited to speech sounds or artificial stimuli such as tones. Both types of harmonic stimuli may engage cortical mechanisms that are dedicated to the processing of speech-specific acoustic features in the human brain. It is therefore unknown whether the flexible and task-dependent cortical processing of category representations observed with speech generalizes to other types of nonvocal sound sources or sound-producing events. Behaviorally, human listeners can reliably identify the material of an impacted sound source or the type of action involved in producing the sound (Lutfi 2007; Hjortkjær and McAdams 2016), but potential task-dependent cortical representations of nonvocal sound sources have not been explored.

The current fMRI study was designed to investigate whether natural sound source information is processed in a task-dependent manner in human cortex. In separate functional MRI scanning sessions, participants had to identify either the material (glass, metal, or wood) or the sound-producing action (striking, rattling, or dropping) from the same set of recorded impact sounds. We analyzed multivoxel patterns of BOLD activity that have been shown to be sensitive to auditory category information undetectable at the single-voxel level (Formisano et al. 2008; Staeren et al. 2009). Classifying both action and material categories based on the sound-evoked response patterns, we asked

whether the different behavioral tasks had any impact on decodable category information. Hypothesizing a cortical bias towards behaviorally relevant information, we predicted a higher decoding accuracy of category information that is task-relevant as opposed to category information that is task-irrelevant. To further examine potential top-down effects of the category tasks on acoustic processing, we characterized the sound stimuli in terms of spectrotemporal features previously shown to be relevant to behavioral identification of material or action information. Using representational similarity analysis (RSA), we examined whether the behavioral tasks might modulate the cortical processing of lower-level sound features or instead operate at the level of abstract category representations irrespective of acoustic content.

Materials and Methods

Participants

A total of 15 healthy subjects (8 females, aged 23–37 years) participated in 2 fMRI sessions with informed written consent. Participants had no history of neurological disorders and reported normal hearing. Experimental procedures were approved by the ethics committees of the Capital Region of Denmark and were conducted in accordance with the Declaration of Helsinki.

Stimuli

We recorded sounds of solid objects made of 1 of 3 types of material (wood, metal, or glass) being manipulated by 1 of 3 types of action (dropping, striking, or rattling) (Fig. 1A). The sounds were recorded at 44.1 kHz sampling rate and 16 kbit/s bit rate in an acoustically shielded room. The levels of the recorded sounds were adjusted by 5 expert listeners so that all stimuli were perceived to be equally loud. For each of the 9 category combinations (3 actions \times 3 materials) we made 9 different sound files of 6 s length, yielding a total of 81 different sound stimuli (with 27 exemplars in each of the category types). The individual 6 s long exemplars were made by concatenating segments of ~1 s corresponding to 6 individual drops or strikes. Drop and strike sounds were made to have irregular impact patterns to avoid any sense of rhythmic repetition.

Detailed behavioral and acoustic analyses of the sound stimuli used in this study are reported in Hjortkjær and McAdams (2016). In that study, laboratory psychophysics suggested that listeners rely on long-term spectral content for identification of material categories and rely on temporal cues for action identification. To examine potential task-dependent cortical processing of acoustic information in the present study, we extracted spectral and temporal features of the sound stimuli using a physiologically plausible computational model of spectrotemporal processing in the early auditory system. The stimuli were first passed through a bandpass filterbank to model cochlear frequency filtering (Glasberg and Moore 1990) followed by half-wave rectification and lowpass filtering at 1 kHz for envelope extraction (Dau et al. 1996). The envelopes at the output of each cochlear filter were compressed to account for basilar membrane nonlinearities (Plack et al. 2008), and passed through a bank of 6 bandpass modulation filters with octave-spaced center frequencies between 2 and 64 Hz (Dau et al. 1997). Finally, the envelope modulation power at the output of each modulation filter was computed (Ewert and Dau 2000).

Experimental Protocol

In 2 fMRI sessions, subjects had to identify either the material or the action categories in the same 81 sound stimuli. The same

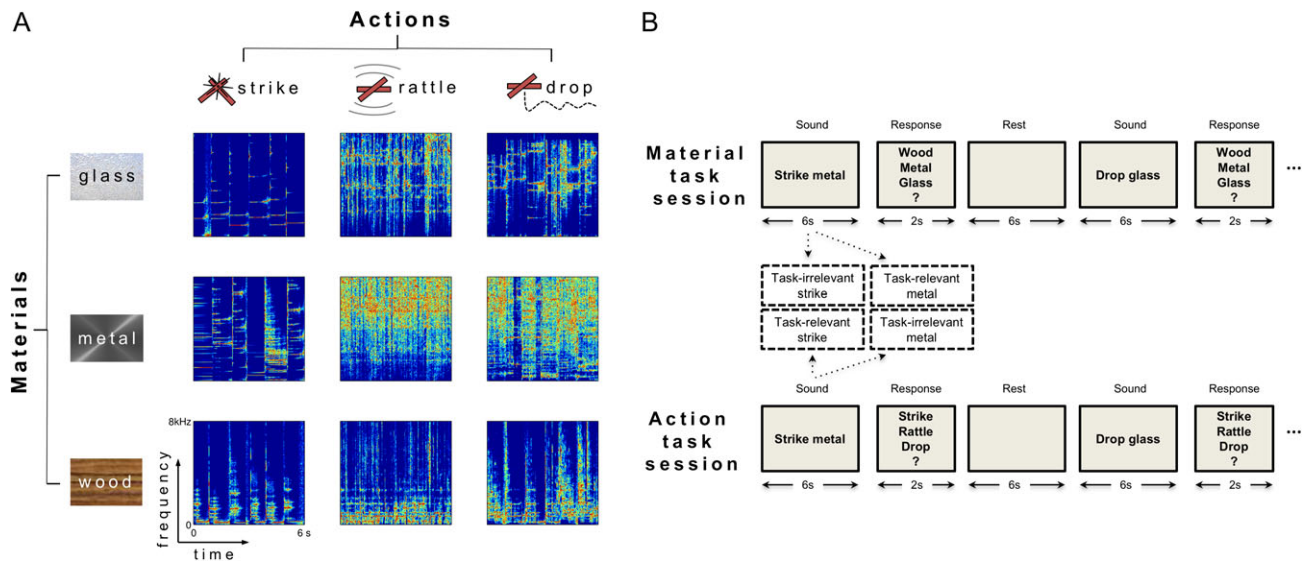


Figure 1. Stimuli and experimental design. (A) Spectrograms of sound stimuli exemplars representing action and material categories. (B) Experimental design. In separate fMRI sessions, subjects identified either the action or the material in the same sound stimuli.

sound stimuli were presented in both sessions via MR-compatible electrodynamic headphones (MR Confon, Magdeburg, Germany) attenuating the background scanner noise. Before the experiment, sound levels were adjusted for each subject with the background scanner noise until the subject reported hearing the sounds clearly. Each presentation trial of 14 s duration began with a 6 s resting baseline followed by the 6 s auditory stimuli and a 2 s response period where subjects indicated which of the 3 categories they had heard via a button press (Fig. 1B). Before each session, subjects performed a short practice session outside the scanner. Subjects were not informed of the alternative category task (materials or actions) until the beginning of the second session. The order of sessions was counterbalanced between subjects.

fMRI Protocol and Image Preprocessing

Brain images were acquired on a 3T Trio Scanner (Siemens, Erlangen). In each of the 2 sessions, 462 volumes of T2*-weighted functional echo planar images (EPI) were acquired (TR = 2490 ms, TE = 30 ms, $3 \times 3 \times 3 \text{ mm}^3$ resolution). Each volume consisted of 42 slices, covering the entire brain except for the lower part of the cerebellum. A high-resolution anatomical image was acquired at the beginning of the scanning sessions using a Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) sequence (TR = 1540 ms, TE = 3.93 ms, $1 \times 1 \times 1 \text{ mm}^3$ resolution). The functional images were spatially aligned and corrected for linear trend. For the univariate GLM analysis, images were normalized to MNI space and spatially smoothed with a Gaussian kernel of 8 mm at full-width half-maximum (FWHM). For the multivoxel pattern analysis, we used spatially aligned but unsmoothed images in native space without further preprocessing. Image preprocessing and GLM analysis was performed using SPM8 (Wellcome Trust Centre for Neuroimaging, University College London, London, UK).

Analysis of Behavioral Data

The accuracy of categorization responses collected during fMRI-scanning was analyzed using a repeated-measures $2 \times 3 \times$

3 analysis of variance (ANOVA) with within-subject factors: task (material/action), material category of the sound stimulus (glass, metal, wood), and action category of the sound stimulus (strike, drop, rattle). The same 3-way ANOVA model was used to test for differences in reaction time. Post hoc paired t-tests with Bonferroni-correction were used to test for differences between individual category combinations. Statistical effects were considered significant at $P < 0.05$. To further quantify perceptual discriminability between categories, we transformed the categorization confusions between all combinations of categories to d' -prime indices of discriminability, defined as $d' = z(\text{hit rate}) - z(\text{false alarm rate})$, where z is the inverse of the cumulative normal distribution.

Univariate fMRI Analysis

For each subject, we fitted a general linear model (GLM) with the onset presentation of each of the 9 category combinations in each session as regressors. The model also included a stick function regressor to capture variance associated with the button-press response. Each regressor was convolved with a canonical hemodynamic response function. Movement parameters estimated from the head motion correction were entered as regressors of no interest. We calculated the contrast of each of the 9 presented category combinations (3 material \times 3 actions) against baseline within each session at the first level. At the group level, we again defined a 3-factorial ($2 \times 3 \times 3$) ANOVA model with tasks (material task/action task), material category of the sound stimulus (glass/metal/wood), and action category of the sound stimulus (strike/drop/rattle) as within-subject factors. Using t-tests, we computed the contrast between the 2 tasks to examine global effects of the type of categorization in terms of regional BOLD activity. To locate regional effects of stimulus exemplars, we computed the contrast between each possible combination of the materials while ignoring the actions, and vice versa (e.g., [glass-strike + glass-drop + glass-rattle] > [metal-strike + metal-drop + metal-rattle + wood-strike + wood-drop + wood-rattle]/2). We then assessed the conjunction between the responses to these different category combinations and also examined

potential interaction effects between tasks and categories. Statistical effects were considered significant at $P < 0.05$ with FWE correction for multiple comparisons at the peak level as implemented in SPM8.

Searchlight Multivoxel Pattern Analysis

We applied a pattern classification approach to identify potential category-sensitive regions in which distributed BOLD responses discriminated between the different sound categories. We used linear kernel support vector machines (SVM) (Vapnik 1995) with default soft margin parameter $c = 1$ (using the LIBSVM package, www.csie.ntu.edu.tw/~cjlin/libsvm). Classification was based on single-trial responses to the sounds. For each trial, we used the mean of the 2 unsmoothed EPI volumes recorded from 5 s after the onset of the sound stimuli to account for the hemodynamic delay. This corresponded to BOLD activity patterns sampled in the period 5–11 s after the onset of the 6-s stimulus. All classifications were carried out in native space to respect individual anatomical variation. We performed the SVM classification to discriminate a given stimulus category presented both during the task-relevant sessions (i.e., classification of the different material categories presented when subjects identified the material, and classification of action categories during action discrimination) as well as the same stimulus categories presented during the task-irrelevant sessions (SVM classification of the material of sound source presented during the action task, and vice versa). For both sessions, we trained separate classifiers to discriminate between each of the 3 pairwise combinations of the action categories (strike vs. rattle, rattle vs. drop, strike vs. drop) and the 3 material categories (glass vs. metal, metal vs. wood, glass vs. wood). We chose to examine pairwise classification rather than discriminating 1 category versus the remaining in order to balance the number of category exemplars to be discriminated. For each of the 3 category combinations, we trained SVM models while leaving out 1 of the 3 categories of the alternative dimension and then tested on the data left out. For instance, a model was trained to discriminate wood versus metal on the drop and strike sounds and then wood versus metal was tested on the rattle sounds. We did this for each of the 3 categories to be held out and then averaged the results of the 3 splits. This classification procedure was used to ensure generalizability of decodable category information, that is, that classifier performance for one category dimension would not be influenced by the presence of the other type of category information. The different pattern classifications were performed inside a local searchlight volume (Kriegeskorte et al. 2006) consisting of a $5 \times 5 \times 5$ voxel cube surrounding each voxel in the brain. We used the univariate GLM to identify voxels in motor, premotor and supplementary motor cortex responding with regional BOLD change to the button press and excluded these voxels from the analysis. Thus, potential processing of action or material categories in these regions was not examined.

The searchlight classification procedure thus resulted in a total of 12 whole-brain discrimination accuracy maps for each participant (2 tasks \times 2 sound category dimensions \times 3 category combinations). To assess the statistical significance of the SVM classification accuracies at the group level, we implemented a fixed effects analysis based on permutation testing (following Stelzer et al. 2013). First, we performed the searchlight classification in each subject as described above using 100 random permutations of the category labels to create 100 “chance”

maps for each category combination. Importantly, the same 100 permutations were used for each searchlight position across the brain so as to preserve spatial correlations and random dependencies between category labels in the chance map (Stelzer et al. 2013). At the group level, we randomly selected one of the 100 chance maps from each subject and averaged these across subjects in MNI space. This bootstrap procedure was repeated 10^5 times to form 10^5 permuted group maps constituting the empirical chance distribution of accuracies at the group level. Voxel-wise P -values for the statistical significance of the true accuracies were computed from this distribution as $(b + 1)/(m + 1)$, where b is the number of random permutations in which the random statistic is greater or equal to the accuracy observed with the correct category labels and m is the total number of permutations (Phipson and Smyth 2010). We implemented a family-wise error (FWE) correction procedure (unlike the FDR correction procedure in Stelzer et al. 2013) to account for the multiple testing problem at the cluster level. We identified clusters in both the true and chance accuracy maps thresholded at $P = 10^{-3}$ and recorded the maximum cluster size in each of the 10^5 chance maps to form a histogram of maximal cluster sizes in the randomly permuted maps (Nichols and Holmes 2002). In the true maps, clusters with sizes exceeding $P < 0.05$ of the permutation distribution were considered statistically significant. We conducted a similar group-level permutation test to assess whether classification accuracies of the same stimulus category combinations were different between the 2 tasks. We randomly permuted the task labels of the accuracies using 10^5 permutations fixed across the brain to form the chance distribution for the null hypothesis of no difference in classification accuracy between the material and action tasks. We computed P -values for the true difference in classification accuracy both from the upper and lower tail of the null distribution to investigate both significantly higher or lower differences in decoding accuracy between the 2 tasks. This enabled us to identify cortical regions that selectively contributed to within-category discrimination for task-relevant as opposed to task-irrelevant category tasks.

Regions of Interest

Apart from the searchlight analysis, multivoxel analyses were performed in 3 anatomically defined regions of interest (ROI) in both hemispheres: Heschl's Gyrus (HG), Planum Temporale (PT), and Inferior Frontal Gyrus (IFG). ROIs were drawn on the high-resolution structural images of each individual subject. For HG, we included both partial and complete duplications of the gyrus if present (Da Costa et al. 2011). The PT was defined as the triangular region extending posteriorly from the transverse sulcus posterior to HG on the supratemporal plane to the posterior-most extent of the Sylvian fissure (Shapleske et al. 1999). The ROI in the IFG comprised pars opercularis, pars triangularis and pars orbitalis of the frontal operculum in both hemispheres. The mean size of each ROI was: left-HG = 93 voxels; right-HG = 103 voxels; left-PT = 266 voxels; right-PT = 242; left-IFG = 823 voxels; right-IFG = 820 voxels.

Relating MVPA Decoding to Behavioral Performance

For each ROI, we computed the multivoxel pattern classification as described above but based on all voxels contained in an ROI. For each subject and each ROI, we then correlated the pattern classification accuracies for each category combination

with the d' measures of perceptual discriminability of the same category combinations.

Representational Similarity Analysis

We used RSA (Kriegeskorte et al. 2008) to relate acoustic features and category structure of the stimuli to multivoxel patterns of BOLD activity in the ROIs. We first computed neural representational dissimilarity matrices (RDMs) from the fMRI data as the pairwise dissimilarity (1 minus the Pearson correlation) between the average multivoxel response pattern to each of the 9 stimulus types by all voxels in an ROI. For each subject, neural RDMs evoked by the same stimuli were computed for both behavioral tasks. Two types of stimulus RDMs were formed to investigate the representation of spectral and temporal stimulus features extracted from the auditory model described above. Spectral RDMs were formed by computing the pairwise dissimilarity between the time-averaged spectra of each sound stimulus at the output of the cochlear filters. Temporal features were quantified in terms of the envelope modulation power spectrum, representing the amount of low-frequency envelope fluctuations created by the different sound-producing actions. For comparison, we also considered the 2 category RDMs that describe which action or material category each stimulus belongs to. For the action category RDM, the dissimilarity between 2 stimuli equals 0 if the stimuli are performed by the same action and 1 otherwise. Similarly, the material category RDM equals 0 for sound stimuli representing the same sound source material and 1 otherwise.

Finally, for each subject and each ROI, we computed Kendall's τ rank correlation coefficient between the lower triangular of the neural RDMs and 2 acoustic feature RDMs as well as between the neural RDMs and 2 category RDMs. The rank correlation was computed excluding the diagonal of the RDMs (which equals 0). Since we expect the acoustic feature RDMs to be correlated with the category RDMs, we also computed the partial Kendall's correlation coefficients for comparison. The partial correlation quantifies the degree of association between neural RDMs and acoustic features with the effect of category features removed, as well as the correlation between the neural and category RDMs with the correlation with acoustic features removed. The correlation values for each subject and each ROI were transformed using the variance-stabilizing Fisher Z-transform. At the group level, we used t-tests to assess whether the correlations between RDMs were significantly different from zero across subjects. To assess the potential influence of the task-relevance on the acoustic features, the correlations for each ROI were analyzed using a repeated-measures ANOVA with factors: task (material/action) and acoustic feature (spectral/temporal) or task (material/action) and dimension of the category RDMs (material/action).

Results

Task Performance

During the fMRI experiment, participants performed both the action and the material categorization tasks with high accuracy (Fig. 2). Categorization performance during the fMRI experiment was highly similar to performance with same sounds measured with listeners in an acoustically controlled laboratory setting (Hjortkjær and McAdams 2016). Mean categorization accuracy across tasks was 96.3% (action categorization task: 98.1%, material categorization task: 94.5%). Subjects occasionally confused some category combinations, in particular metal and glass sounds, consistent with previous behavioral results (Hjortkjær

and McAdams 2016; Giordano and McAdams 2006), but we found no statistically significant differences between the 3 material categories ($F_{2,14} = 0.42$, $P > 0.7$), between action categories ($F_{2,14} = 0.36$, $P > 0.7$), or between the 2 tasks ($F_{1,14} = 4.08$, $P > 0.06$). Mean response times were also not significantly different between the material categories ($F_{2,14} = 0.93$, $P > 0.41$), between action categories ($F_{2,14} = 0.26$, $P > 0.77$), or between the 2 tasks ($F_{1,14} = 0.29$, $P > 0.6$).

Univariate BOLD Responses to Sound Source Categories

Univariate whole-brain analysis of stimulus effects yielded no systematic differences between categories and no significant effect of the behavioral tasks in terms of the regional BOLD activity. We also did not find any interaction between the presented sound categories and the tasks in the voxel-wise analyses of the data.

Multivoxel Pattern Decoding of Sound Source Categories

In contrast to the voxel-wise analysis, searchlight multivoxel pattern classification identified category-sensitive regions in superior temporal and frontal cortices. Figure 3 shows cortical areas in which BOLD activity patterns discriminated the individual sound categories within a given category dimension. In particular, auditory cortical regions, including HG and the PT region in both hemispheres, discriminated both material and action categories consistently across different category combinations. Some combinations of material categories were also discriminated in anterior superior temporal and inferior frontal cortices although less consistently than in auditory cortex.

To investigate the impact of the behavioral tasks on cortical category processing, we compared multivoxel classification performance during the 2 tasks. Specifically, we assessed differences in the accuracy with which the same sound categories could be decoded based on the response patterns evoked by the same sound stimuli during the task-relevant and the task-irrelevant sessions (Fig. 4). For both the material and action task, we found that decoding success depended on the behavioral context introduced by the task. Response patterns in auditory cortex, with a peak contrast effect in PT bilaterally, discriminated action categories with higher accuracy in sessions where participants categorized the actions, relative to sessions in which subjects categorized the materials of the same sound stimuli. Examining each of the category combinations individually (Fig. 4B), we found that voxels in the PT region discriminated action categories above chance level only when subjects performed the action categorization task, and not during the material categorization task. A similar task dependency was not found for the material categories in auditory cortex. For the material categories, response patterns in the middle and inferior frontal cortex discriminated material categories with higher accuracy when participants categorized the material. A smaller region in the opercular part of the IFG also discriminated action categories with higher accuracy when the action information was task-relevant.

Task-Modulation of Acoustic Features in Auditory Cortex

Whole-brain searchlight analysis identified task-modulated auditory cortical regions but with low spatial accuracy, including voxels in both lower and higher-order auditory cortex. To

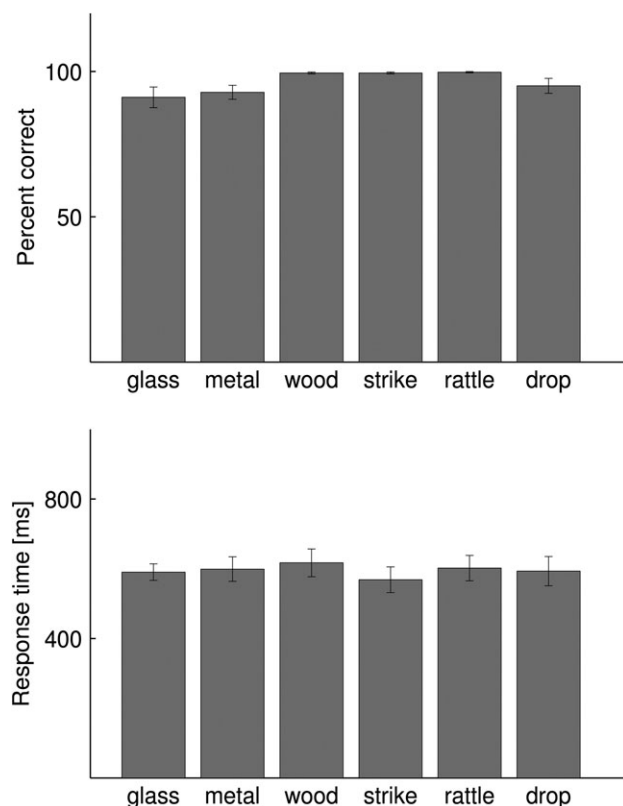


Figure 2. Behavioral data. Percent correct and reaction times for identification of the task-relevant category during the fMRI experiment. For both accuracy and reaction times, there were no significant differences neither between the 2 category dimensions nor within the action or material categories. Error bars denote \pm S.E.M.

investigate potential task-dependent representations of acoustic information, we performed additional multivoxel analyses based on voxels restricted to anatomically defined ROIs. Using RSA (Fig. 5A), we characterized the sound stimuli in terms of their pairwise similarity in spectral content or temporal modulations, previously shown to be behaviorally relevant to material and action categorization, respectively (Hjortkjaer and McAdams 2016). We compared these acoustic similarity representations with the similarity of neural response patterns evoked by each sound during the 2 tasks in a particular ROI. The right panel in Figure 5B shows the rank correlation between neural and acoustic similarity matrices, measured with Kendall's τ , for the 2 tasks. In HG, containing the primary auditory fields, we found that the similarity of neural response patterns correlated significantly with spectral similarity, which is expected given the tonotopic organization of PAC (Romani et al. 1975). Moreover, spectral similarity in HG was significantly modulated by the categorization task ($F_{1,14} = 5.11$, $P < 0.04$). Compared with the action task, the correlation with the spectral similarity between the sound stimuli was enhanced during the material task. In the PT region, the similarity of response patterns correlated significantly with the similarity of both acoustic features, producing a significant interaction between tasks and acoustic features ($F_{1,14} = 4.6$, $P < 0.05$). In the PT, the material task enhanced the correlation for the spectral feature, while the action task enhanced the correlation for the temporal modulations in the sound stimuli. Acoustic features were not correlated with the similarity of response patterns in IFG.

For comparison, we also related ROI response patterns to category-level RDMs describing which material or action category each sound stimulus belongs to (Fig. 5B, left panel). Across auditory cortex, sounds produced by the same action elicited similar response patterns, with an enhanced correlation for the task-relevant categories in the PT region as also suggested by our searchlight analysis (task \times category interaction in PT: $F_{1,14} = 18.11$, $P < 0.001$). The same sound source materials elicited similar activity patterns in the IFG during the material task, but they were negatively correlated during the action task (task \times category interaction in IFG: $F_{1,14} = 31.55$, $P < 0.001$), again supporting the results of the searchlight classification.

Decoding Success Predicts Behavioral Performance

To examine whether multivoxel classification accuracy predicted behavioral performance, we correlated decoding accuracies in each ROI with behavioral measures of category discriminability, as measured by d' . During the task-relevant sessions, we found that behavioral discriminability correlated positively with decoding accuracy in HG ($r = 0.16$, $P < 0.03$) and PT ($r = 0.29$, $P < 0.001$) but not in IFG ($r = -0.08$, $P > 0.23$). In effect, category combinations that were better discriminated behaviorally elicited more discriminable response patterns in auditory cortex. Interestingly, we also found that decoding accuracies for the task-irrelevant categories were negatively correlated with behavioral performance in auditory cortex (HG: $r = -0.34$, $P < 0.001$; PT: $r = -0.44$, $P < 0.001$). As a result, the difference in decoding accuracy between the task-relevant and task-irrelevant categories strongly predicted behavioral discriminability in auditory cortex (HG: $r = 0.36$, $P < 0.001$; PT: $r = 0.45$, $P < 0.001$) but not in IFG ($r = 0.11$, $P > 0.11$).

Discussion

Our results suggest that the human cortex extracts detailed category information from natural sound sources in a task-dependent manner. Analyzing local BOLD response patterns, we identified regions in the temporal auditory cortex and frontal cortex that discriminated information about sound source materials and sound-producing actions. The degree to which cortical response patterns in these regions could be used to discriminate sound source categories was found to depend on task relevance of the category information, as well as on perceptual discriminability of the different category combinations.

Category-Selective Cortical Representations

Consistent discrimination of within-category information for both the action and the material dimension was observed across auditory cortex (Fig. 3). Detailed sound source information was only decodable with pattern analysis of locally distributed cortical activity, and not by differences in single-voxel responses. For instance, we did not find single voxels responding with different BOLD amplitudes to sound sources made of wood, glass or metal or to sounds produced by striking, rattling or dropping. Previous imaging studies of real-world environmental sounds have reported regional cortical activations that differentiate coarse object categories, such as living versus non-living sound sources (Lewis et al. 2004, 2005, 2011; Engel et al. 2009; De Lucia et al. 2009; Doehrmann et al. 2008; Leaver and Rauschecker 2010; Giordano et al. 2012). Cortical representations of detailed within-category information have so far

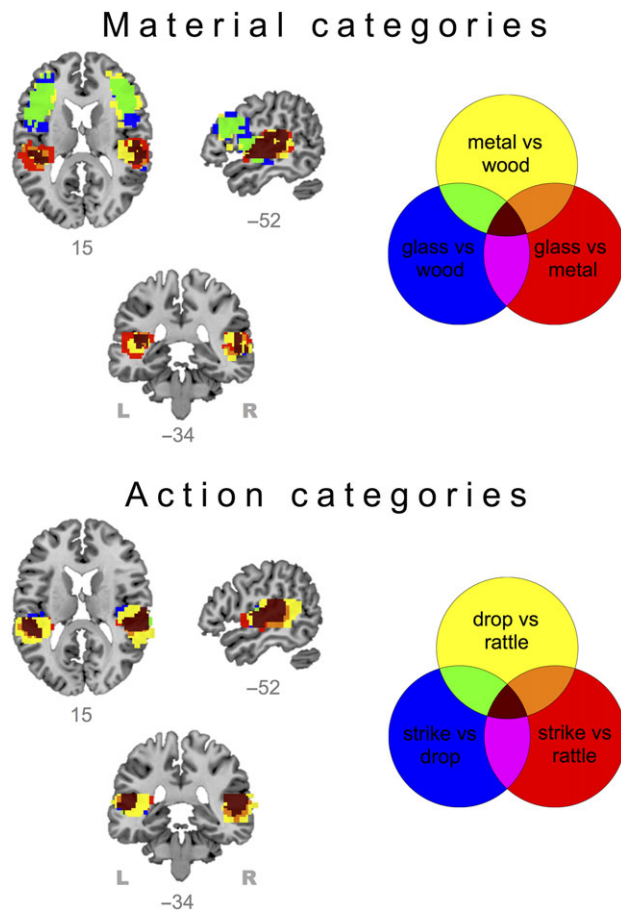


Figure 3. Cortical regions sensitive to sound source category information. Group-level discriminative maps obtained with a multivoxel classification searchlight showing the anatomical overlap of regions discriminating between different combinations of action (below) and material (above) categories ($P < 0.05$, FWE corrected).

mainly been demonstrated for speech sounds (Formisano et al. 2008; Kilian-Hütten et al. 2011; Bonte et al. 2014). Speech is indeed a primary exponent of robust category perception in humans, but the ability of the primate brain to extract detailed category information from sound sources presumably generalizes beyond speech processing. Our results showing robust decoding of detailed sound source information concur with behavioral evidence showing that listeners are able to infer material and action information from impacted sound sources (Warren and Verbrugge 1984; Giordano and McAdams 2006; McAdams et al. 2010; Hjortkjær and McAdams 2016). We therefore suggest that this type of sound source information may be viewed as analogous to visual object categories and constitutes an important stimulus feature for probing category processing in auditory cortex.

Task-Dependency of Category Information

Multivariate pattern classification allowed us to characterize differences in the cortical representation of within-category information during the different behavioral tasks. In contrast to the task-dependent differences in pattern decoding, the univariate subtraction between regional BOLD levels during the action-task and material-task sessions did not reveal any significant task-specific differences. This suggests that the more accurate decoding of task-relevant category information cannot be explained by

a simple gain in neural activity due to differences in attention or task difficulty (Hillyard et al. 1998; Petkov et al. 2004). An influence of category-level tasks on distributed category representations has previously been demonstrated for speech sounds in superior temporal cortex (Kilian-Hütten et al. 2011; Bonte et al. 2014). Task-effects have been reported in STG/STC regions that represent speech sounds categorically, possibly by integrating acoustic content over time and frequency to form broader spectrotemporal response properties (Mesgarani et al. 2014). Similar to our findings, task-dependent pattern decoding of speaker versus vowel information (Bonte et al. 2014) or of the perceptual interpretation of syllables (Kilian-Hütten et al. 2011) with identical acoustic input has also been reported in HG. The present results suggest that task-dependent category processing in both early and higher-order auditory cortex generalizes beyond speech categories.

Role of Spectral and Temporal Features

Behavioral psychophysics of the sound stimuli used in this study previously uncovered the acoustical features that listeners use to identify action and material categories when these category dimensions co-vary (Hjortkjær and McAdams 2016). The behavioral evidence suggested that listeners rely mainly on long-term spectral information to identify materials across the acoustic variation introduced by the different actions. Action recognition across the materials, on the other hand, relied on the temporal pattern of amplitude variations created by the different actions and actions could be recognized even without spectral information (Warren and Verbrugge 1984; Hjortkjær and McAdams 2016). Using representational similarity analysis, our current results suggested that the representation of temporal and spectral information in early and higher-order auditory cortex was modulated in favor of task-relevant category information. Compared with the action task, material identification enhanced the correlation between the spectral similarity of the stimuli and the similarity of the multivoxel responses in both HG and PT. In PT, but not in HG, temporal modulations relevant to action recognition was enhanced by the action task. Previous studies have also reported sensitivity to spectral content in HG and anterior parts of the auditory cortex (Warren et al. 2005; Zatorre and Belin 2001) and encoding of spectrotemporal modulations with higher temporal detail in more posterior AC regions (Santoro et al. 2014; Giordano et al. 2012; Kuśmierek and Rauschecker 2014), but our results further suggest that category-level tasks may modulate the processing of spectral and temporal information to optimize the representation of relevant object categories.

Role of Behavioral Category Discriminability

Apart from the task dependency of category representations, decoding accuracy of the individual category combinations also predicted participant's discrimination accuracy. For the task-relevant categories, the discrimination performance correlated positively with decoding accuracy in HG and PT. Interestingly, we also found that behavioral performance was negatively correlated with decoding accuracy for the task-irrelevant categories in auditory cortex. We note that this opposite direction of correlation was found despite the fact that we used a classification scheme that ensured independent decoding of task-relevant and task-irrelevant category information. A previous study on speech sounds (Bonte et al. 2014) similarly reported a positive correlation between behavioral discrimination performance and speaker decoding accuracy in voice-selective

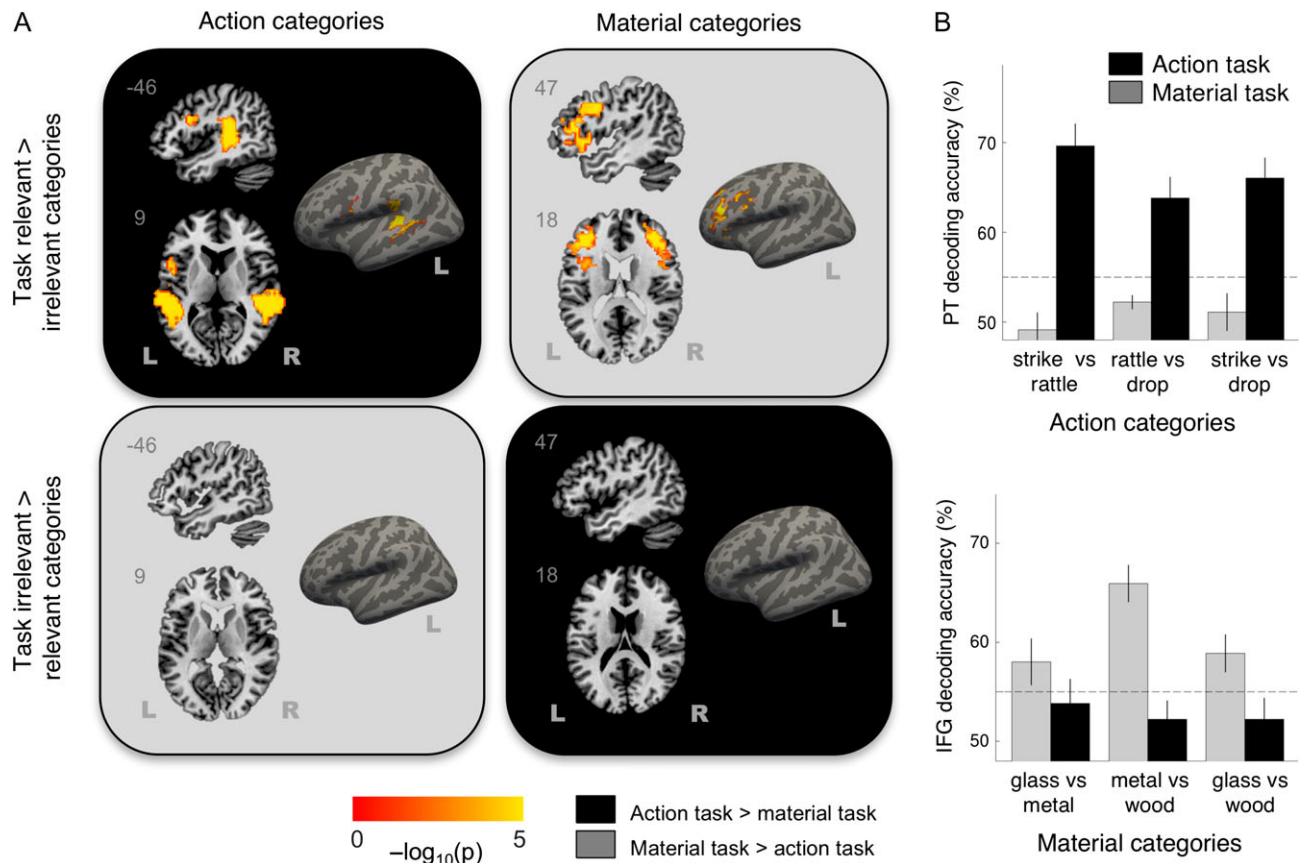


Figure 4. Task effects on neural category representations. (A) Contrast maps showing cortical regions that encode sound source category information differently during relevant versus irrelevant tasks. Response patterns in auditory cortex, including the planum temporale (PT) region, discriminated the sound-producing action of the sound stimuli with higher accuracy during the action categorization task. In the IFG, category information about the material of the sound source was better discriminated during the relevant material task. No regions showed an enhanced encoding of task-irrelevant category information. The maps show logarithmically transformed P -values ($-\log_{10}(P)$) for significant differences in decoding accuracy between the tasks determined with permutation tests ($P < 0.05$, FWE corrected). (B) Decoding accuracies for each category combination in the right PT for the action categories (center voxel at $x = 66$, $y = -34$, $z = 19$ mm) and in the left IFG for the materials ($x = -45$, $y = 20$, $z = 2$ mm). Error bars denote \pm S.E.M. Horizontal lines represent the statistical chance level ($P < 0.001$, uncorrected).

regions within the left superior temporal cortex but not in HG. In our study, we found that category decoding accuracy correlated with behavioral performance in the PT but also at the level of the HG.

The relationship between perceptual discriminability and category decoding may also explain why a task-modulation in auditory cortex was only observed for the action categories (Fig. 4). For materials, previous psychoacoustic studies have indicated that confusions between glass and metal sounds occur because of the spectral overlap between these broadband sounds (Giordano and McAdams 2006; McAdams et al. 2010; Hjortkjær and McAdams 2016). Material categorization may thus have presented a perceptually more demanding discrimination task in spite of high response accuracy. The resulting lower difference in metal-glass classification accuracies between the task-relevant and task-irrelevant responses may thus explain the lack of a consistent task-effect for materials across category combinations in auditory cortex.

Auditory Categorization in Auditory and Prefrontal Cortex

Task effects on decodable category information was observed in both auditory and prefrontal cortex. Category-related activity patterns in prefrontal cortex could suggest processing related

to decision-making involved in the task. A network comprising the auditory cortex and the prefrontal cortex is consistently reported in studies on context-dependent sound category processing (Russ et al. 2008; Fritz et al. 2010; David et al. 2012; Bonte et al. 2014). Consistent with the present results, the prefrontal cortex has been shown to encode the category membership of an auditory stimulus depending on its behavioral meaning rather than on category-specific acoustic features (Cohen et al. 2006; Gifford et al. 2005; Romanski et al. 2005; Russ et al. 2007; Lee et al. 2009). The prefrontal cortex has been also shown to trigger receptive field changes in auditory cortex via direct prefrontal-AC pathways (Winkowski et al. 2013, 2017). According to the notion of “behavioral gating,” the prefrontal cortex modulates feature representations in AC to achieve invariance along a particular task-relevant category dimension of the auditory stimulus (Fritz et al. 2010; David et al. 2012; Atiani et al. 2014). This hypothesis is compatible with our results showing task-modulated category information unrelated to acoustic information in the prefrontal cortex as well as modulation of target-specific acoustic features in the AC. Greater task effects in AC for the action categories could be related to a higher degree of discriminability of the action information, as discussed above. This is also consistent with single-unit results showing that online changes in spectrotemporal tuning properties in AC scale with behavioral performance or

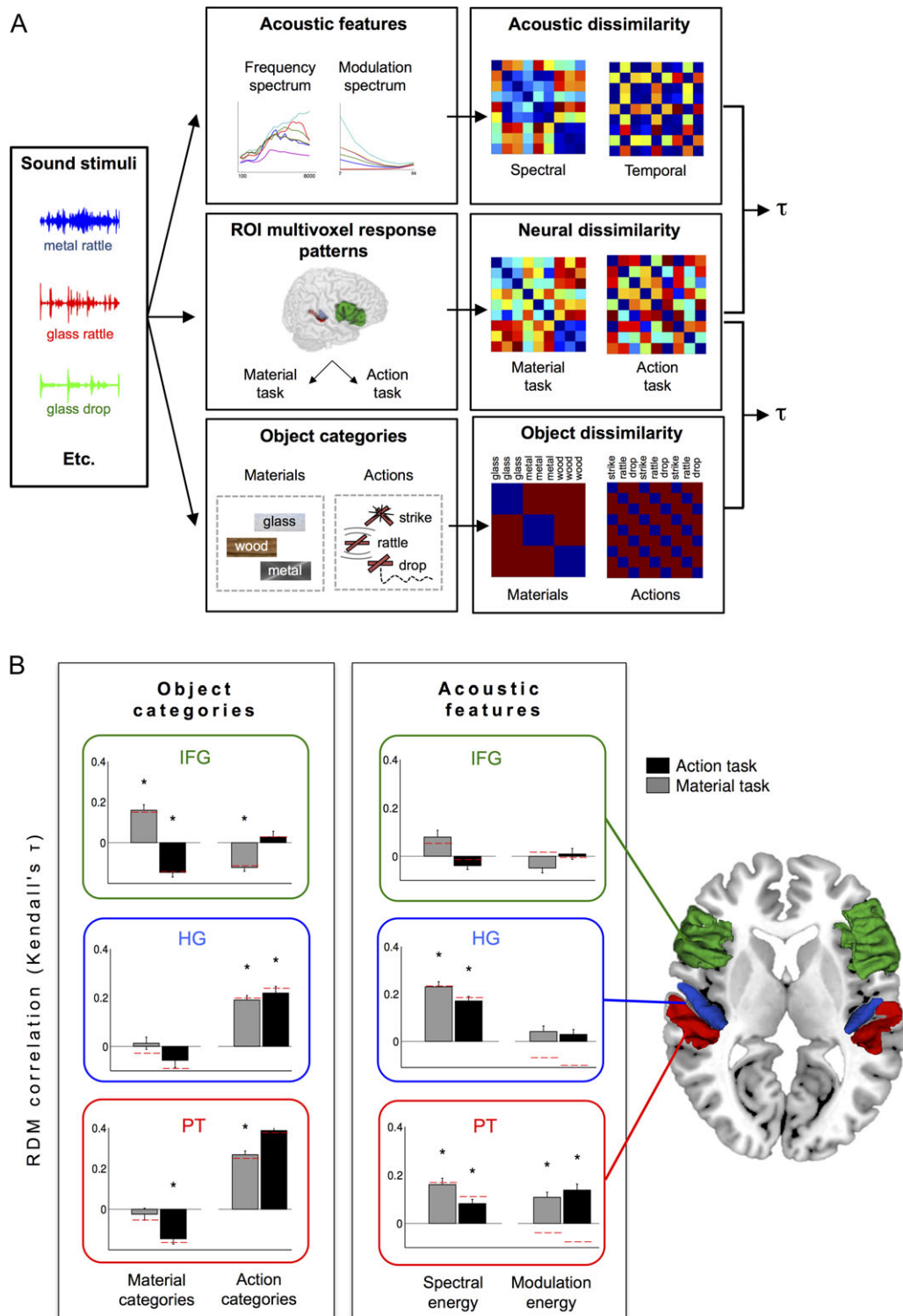


Figure 5. Representational similarity analysis (RSA) in anatomically defined regions-of-interest. (A) RSA analysis comparing representational dissimilarity matrices (RDMs) derived from spectral and temporal acoustic features of the presented sound stimuli, ROI multivoxel response patterns during the 2 tasks, and the object categories of the sound stimuli. (B) The RDM correlation between response patterns in auditory cortex and acoustic features (right) and category structure (left) is modulated by the behavioral tasks. The RDM correlation between multivoxel activity patterns in Heschl's gyrus and spectral content of the sound stimuli is enhanced during material identification known to rely on spectral information. Representation of temporal modulations is enhanced in favor of the relevant action categorization task in the planum temporale region. Red stippled lines indicate the partial correlation of the acoustic features when removing the effect of the object categories, and the partial correlation of the object categories with the effect of the acoustic features removed. The colored ROIs indicate the spatial extent of the analyzed regions on the MNI template brain. Error bars denote \pm S.E.M. * $P < 0.05$.

task difficulty (Atiani et al. 2009). Less discriminable material categories, on the other hand, may engage abstract category representations in the prefrontal cortex that are not directly related to acoustic features.

The prefrontal cortex may also be more closely related to category processing that enables a specific motor output. The presence or absence of an overt motor response during auditory categorization has been found to induce different types of change in receptive fields of single neurons in A1. David et al. (2012) showed that an avoidance task increased receptive field tuning at the frequency of the target sound, whereas an approach task decreased the response. The prefrontal cortex may trigger such adaptive responses by signaling the behavioral value of the stimulus to downstream motor areas to initiate or inhibit motor responses (Fritz et al. 2010; David et al. 2012). In the present paradigm, subjects were asked to indicate their response via a button press and it is possible that the nature of the adaptive changes in AC feature encoding is contingent on the presence of a motor response. However, although the motor association may lead to either enhanced or suppressed spectrotemporal tuning, both types of change amplifies discriminability between the sound classes to be categorized in the auditory task (David et al. 2012). Decoding methods are blind to such changes in the feature representation since they measure the resulting changes in category discriminability at the population level. Task-dependent differences in decoding accuracy observed in the current study could thus be caused by different types of change in the feature encoding interacting with the motor output. However, our current paradigm does not allow us to disentangle the influence of the motor response and potential auditory-motor associations in the prefrontal-auditory cortical network remain to be explored.

Conclusions

Our results point to a network comprising the auditory and prefrontal cortices that support dynamic categorization of everyday sounds during auditory behavior. Studying listeners engaged in categorization of natural nonspeech sounds, we show that information about sound source materials and sound-producing actions is represented with a bias depending on their task relevance. Our analyses suggested that behavior modulates response patterns at the level of the auditory cortex to enhance spectrotemporal features that discriminate between task-relevant categories. Response patterns in the frontal cortex, on the other hand, discriminated task-relevant categories abstracted from the acoustic properties of the sound stimulus.

Funding

Lundbeck Foundation (Grant of Excellence “Control of Action” R59 A5399 to H.R.S.), the Independent Research Fund Denmark (J.H.), the Oticon Foundation (CHES: Oticon Centre of Excellence for Hearing and Speech Sciences) (J.H.), EU H2020-ICT (Grant number 644732, COCOHA: Cognitive Control of a Hearing Aid) (J.H.), and the Swiss National Science Foundation (T.K.).

Notes

We would like to thank the anonymous reviewers for insightful comments that led to important new analyses and interpretation of results. *Conflict of Interest:* The authors declare no competing financial interests.

References

- Ahveninen J, Hämäläinen M, Jääskeläinen IP, Ahlfors SP, Huang S, Lin FH, Rai T, Sams M, Vasios CE, Belliveau JW. 2011. Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *Proc Natl Acad Sci USA*. 108:4182–4187.
- Ahveninen J, Jääskeläinen IP, Raj T, Bonmassar G, Devore S, Hämäläinen M, Levänen S, Lin FH, Sams M, Shinn-Cunningham BG, et al. 2006. Task-modulated “what” and “where” pathways in human auditory cortex. *Proc Natl Acad Sci USA*. 103:14608–14613.
- Alain C, Arnott SR. 2000. Selectively attending to auditory objects. *Front Biosci*. 5:D202–D212.
- Andics A, McQueen JM, Petersson KM, Gál V, Rudas G, Vidnyánszky Z. 2010. Neural mechanisms for voice recognition. *Neuroimage*. 52:1528–1540.
- Atiani S, David SV, Elgueda D, Locastro M, Radtke-Schuller S, Shamma SA, Fritz JB. 2014. Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron*. 82:486–499.
- Atiani S, Elhilali M, David SV, Fritz JB, Shamma SA. 2009. Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron*. 61:467–480.
- Bizley JK, Cohen YE. 2013. The what, where and how of auditory-object perception. *Nat Rev Neurosci*. 14:693–707.
- Bonte M, Hausfeld L, Scharke W, Valente G, Formisano E. 2014. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J Neurosci*. 34:4548–4557.
- Brechmann A, Scheich H. 2004. Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cereb Cortex*. 15:578–587.
- Cohen YE, Hauser MD, Russ BE. 2006. Spontaneous processing of abstract categorical information in the ventrolateral prefrontal cortex. *Biol Lett*. 2:261–265.
- Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M. 2011. Human primary auditory cortex follows the shape of Heschl’s gyrus. *J Neurosci*. 31:14067–14075.
- Da Costa S, van der Zwaag W, Miller LM, Clarke S, Saenz M. 2013. Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex. *J Neurosci*. 33:1858–1863.
- Dau T, Kollmeier B, Kohlrausch A. 1997. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *J Acoust Soc Am*. 102:2892–2905.
- Dau T, Puschel D, Kohlrausch A. 1996. A quantitative model of the ‘effective’ signal processing in the auditory system. I. Model structure. *J Acoust Soc Am*. 99:3615–3622.
- David SV, Fritz JB, Shamma SA. 2012. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc Natl Acad Sci USA*. 109:2144–2149.
- Degerman A, Rinne T, Salmi J, Salonen O, Alho K. 2006. Selective attention to sound location or pitch studied with fMRI. *Brain Res*. 1077:123–134.
- De Lucia M, Cocchi L, Martuzzi R, Meuli RA, Clarke S, Murray MM. 2009. Perceptual and semantic contributions to repetition priming of environmental sounds. *Cereb Cortex*. 20:1676–1684.
- Desai R, Liebenthal E, Waldron E, Binder JR. 2008. Left posterior temporal regions are sensitive to auditory categorization. *J Cogn Neurosci*. 20:1174–1188.

- Doehrmann O, Naumer MJ, Volz S, Kaiser J, Altmann CF. 2008. Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia*. 46:2776–2786.
- Engel LR, Frum C, Puce A, Walker NA, Lewis JW. 2009. Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage*. 47:1778–1791.
- Ewert SD, Dau T. 2000. Characterizing frequency selectivity for envelope fluctuations. *J Acoust Soc Am*. 108:1181–1196.
- Formisano E, De Martino F, Bonte M, Goebel R. 2008. “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science*. 322:970–973.
- Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA. 2010. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat Neurosci*. 13:1011.
- Fritz J, Shamma S, Elhilali M, Klein D. 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci*. 6:1216–1223.
- Gaver WW. 1993. What in the world do we hear?: An ecological approach to auditory event perception. *Ecol Psychol*. 5:1–29.
- Gifford GW, MacLean KA, Hauser MD, Cohen YE. 2005. The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. *J Cogn Neurosci*. 17:1471–1482.
- Giordano B, McAdams S. 2006. Material identification of real impact sounds: effects of size variation in steel, glass, wood, and plexiglass plates. *J Acoust Soc Am*. 119:1171–1181.
- Giordano B, McAdams S, Zatorre RJ, Kriegeskorte N, Belin P. 2012. Abstract encoding of auditory objects in cortical activity patterns. *Cereb Cortex*. 23:2025–2037.
- Glasberg BR, Moore BC. 1990. Derivation of auditory filter shapes from notched-noise data. *Hear Res*. 47:103–138.
- Griffiths TD, Warren JD. 2004. What is an auditory object? *Nat Rev Neurosci*. 5:887–892.
- Hillyard SA, Hink RF, Schwent VL, Picton TW. 1973. Electrical signs of selective attention in the human brain. *Science*. 182:177–180.
- Hillyard SA, Vogel EK, Luck SJ. 1998. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philos Trans R Soc Lond B Biol Sci*. 353:1257–1270.
- Hjortkjær J, McAdams S. 2016. Spectral and temporal cues for perception of material and action categories in impacted sound sources. *J Acoust Soc Am*. 140:409–420.
- Jäncke L, Mirzazade S, Shah NJ. 1999. Attention modulates activity in the primary and the secondary auditory cortex. *Neurosci Lett*. 266:125–128.
- Kauramäki J, Jääskeläinen IP, Sams M. 2007. Selective attention increases both gain and feature selectivity of the human auditory cortex. *PLoS One*. 2:e909.
- Kilian-Hütten N, Valente G, Vroomen J, Formisano E. 2011. Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J Neurosci*. 31:1715–1720.
- Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc Natl Acad Sci USA*. 103:3863–3868.
- Kriegeskorte N, Mur M, Bandettini PA. 2008. Representational similarity analysis-connecting the branches of systems neuroscience. *Front Syst Neurosci*. 2:4.
- Kuśmieriek P, Rauschecker JP. 2014. Selectivity for space and time in early areas of the auditory dorsal stream in the rhesus monkey. *J Neurophysiol*. 111:1671–1685.
- Leaver AM, Rauschecker JP. 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J Neurosci*. 30:7604–7612.
- Lee JH, Russ BE, Orr LE, Cohen Y. 2009. Prefrontal activity predicts monkeys’ decisions during an auditory category task. *Front Integr Neurosci*. 3:16.
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA. 2005. Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci*. 25:5148–5158.
- Lewis JW, Talkington WJ, Puce A, Engel LR, Frum C. 2011. Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *J Cogn Neurosci*. 23:2079–2101.
- Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA. 2004. Human brain regions involved in recognizing environmental sounds. *Cereb Cortex*. 14:1008–1021.
- Lutfi RA. 2007. Human sound source identification. In: Yost WA, Popper AN, editors. *Auditory perception of sound sources*. New York: Springer. p. 13–42.
- McAdams S, Roussarie V, Chaigne A, Giordano BL. 2010. The psychomechanics of simulated sound sources: material properties of impacted thin plates. *J Acoust Soc Am*. 128:1401–1413.
- Mesgarani N, Chang EF. 2012. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*. 485:233–236.
- Mesgarani N, Cheung C, Johnson K, Chang EF. 2014. Phonetic feature encoding in human superior temporal gyrus. *Science*. 343:1006–1010.
- Nichols TE, Holmes AP. 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*. 15:1–25.
- O’Connell MN, Barczak A, Schroeder CE, Lakatos P. 2014. Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. *J Neurosci*. 34:16496–16508.
- Paltoglou AE, Sumner CJ, Hall DA. 2009. Examining the role of frequency specificity in the enhancement and suppression of human cortical activity by auditory selective attention. *Hear Res*. 257:106–118.
- Paltoglou AE, Sumner CJ, Hall DA. 2011. Mapping feature-sensitivity and attentional modulation in human auditory cortex with functional magnetic resonance imaging. *Eur J Neurosci*. 33:1733–1741.
- Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL. 2004. Attentional modulation of human auditory cortex. *Nat Neurosci*. 7:658–663.
- Phipson B, Smyth GK. 2010. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat Appl Genet Mol Biol*. 9. doi: 10.2202/1544-6115.1585.
- Pizzamiglio L, Aprile T, Spitoni G, Pitzalis S, Bates E, D’amico S, Di Russo F. 2005. Separate neural systems for processing action- or non-action-related sounds. *Neuroimage*. 24:852–861.
- Plack CJ, Oxenham AJ, Simonson AM, O’Hanlon CG, Draga V, Arifanto D. 2008. Estimates of compression at low and high frequencies using masking additivity in normal and impaired ears. *J Acoust Soc Am*. 123:4321–4330.
- Riecke L, Peters JC, Valente G, Kemper VG, Formisano E, Sorger B. 2016. Frequency-selective attention in auditory scenes recruits frequency representations throughout human superior temporal cortex. *Cereb Cortex*. 27:3002–3014.
- Romani GL, Williamson SJ, Kaufman L. 1975. Tonotopic organization of the human auditory cortex. *Psychiatry*. 132:650.

- Romanski LM, Averbeck BB, Diltz M. 2005. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J Neurophysiol.* 93:734–747.
- Russ BE, Orr LE, Cohen YE. 2008. Prefrontal neurons predict choices during an auditory same-different task. *Curr Biol.* 18:1483–1488.
- Russ BE, Lee YS, Cohen YE. 2007. Neural and behavioral correlates of auditory categorization. *Hear Res.* 229:204–212.
- Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, Formisano E. 2014. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput Biol.* 10:e1003412.
- Shapleske J, Rossell SL, Woodruff PW, David AS. 1999. The plenum temporale: a systematic, quantitative review of its structural, functional and clinical significance. *Brain Res Rev.* 29:26–49.
- Staeren N, Renvall H, De Martino F, Goebel R, Formisano E. 2009. Sound categories are represented as distributed patterns in the human auditory cortex. *Curr Biol.* 19:498–502.
- Stelzer J, Chen Y, Turner R. 2013. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage.* 65:69–82.
- Vapnik VN. 1995. The nature of statistical learning theory. Berlin: Springer.
- von Kriegstein K, Smith DRR, Patterson RD, Kiebel SJ, Griffiths TD. 2010. How the human brain recognizes speech in the context of changing speakers. *J Neurosci.* 30:629–638.
- Warren JD, Jennings AR, Griffiths TD. 2005. Analysis of the spectral envelope of sounds by the human brain. *Neuroimage.* 24:1052–1057.
- Warren WH, Verbrugge RR. 1984. Auditory perception of breaking and bouncing event: a case study. *J Exp Psychol.* 10:704–712.
- Winkowski DE, Bandyopadhyay S, Shamma SA, Kanold PO. 2013. Frontal cortex activation causes rapid plasticity of auditory cortical processing. *J Neurosci.* 33:18134–18148.
- Winkowski DE, Nagode DA, Donaldson KJ, Yin P, Shamma SA, Fritz JB, Kanold PO. 2017. Orbitofrontal cortex neurons respond to sound and activate primary auditory cortex neurons. *Cereb Cortex.* 1–12. doi: 10.1093/cercor/bhw409. [Epub ahead of print]
- Zatorre RJ, Belin P. 2001. Spectral and temporal processing in human auditory cortex. *Cereb Cortex.* 11:946–953.
- Zatorre RJ, Bouffard M, Belin P. 2004. Sensitivity to auditory object features in human temporal neocortex. *J Neurosci.* 24:3637–3642.